Developments in Medical Elementology and Spectral Diagnostics of Disease Via Chemometrics and Machine Learning Assisted Trace Spectroanalytics and Imaging Towards Applications in Nanomedicine

Angeyo[1] H. A., Kaduki[1] K. A., Bulimo[2] D. W. Dehayem-Massop[1] A.

[1]Department of Physics, University of Nairobi, Kenya.
[2]Department of Biochemistry, University of Nairobi, Kenya

MOTIVATION

The *medical elementology and spectral diagnostics of disease* research line in our group involves method development in spectroanalytical and hyperspectral imaging (in vivo and ex vivo) of pathogens and trace chemicals (especially elements) and their speciation in malarial and cancer body tissues and fluids. The goal is to identify and quantify the pathogens and trace and ultra-trace chemicals and their speciation, and to elucidate their role in biopathological activity in cancer and malaria, including modelling their evolution with disease (dynamic analysis and imaging) in the body fluids and tissue.

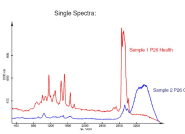

Fig 1. Raman spectra for normal and malignant breast tissue.

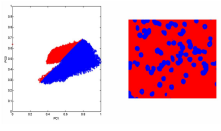

Fig 2. Ppixels of red blood cells for a transmission image (colour coded-blue) delineated in score space and highlighted in image space.

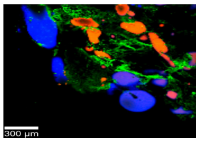

Fig 3. Raman image of normal breast tissue.

INTRODUCTION

The chemical composition of body fluids and tissue are correlated to the state of health. The concentration level and speciation alterations, as well as correlations of trace chemicals in body tissues and fluids can thus be used as parameters for disease diagnosis, as diseases lead to chemical and structural changes in the body that alter the spectral and image characteristics that may be obtained with spectrocopoic/ imaging of tissue and fliud samples: Each tissue and fliud has a characteristic biochemical composition that alters in response to anatomical and pathological stimuli. We are focused on cancer and malaria as there is a connection between Epstein-Barr virus, falciparum malaria, and some forms of cancer in Africa. It has also been found that similar drugs are effective in curing both.

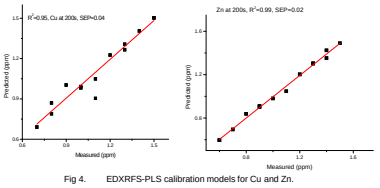

Fig 4. EDXRFS-PLS calibration models for Cu and Zn.

Table 1. TXRF trace metal analysis of dog cancer tissues

Tissue Type	Present	Dominant
Cancerous tissues	Ni, Ca, Br, Cu, Se, Mn, Zn, S, P, K, Fe, Rb, Cl	Ca, Br, Cu, Se, Mn, Zn, Fe
Rhabdomyo sarcoma 1	Cu	Cu
Rhabdomyo sarcoma 2	Ni, Ca, Br, Cu	Cu, Br
Lymphoma 1	Cl, Fe, Rb	Se, Fe
Lymphoma 2	Mn, Zn, S, P, K	Se, K

Alterations in trace element homeostasis are also associated with pathologic processes. For example, malaria infection causes structural and molecular damage to body cells and alters the levels of trace elements in blood. Changes in the concentration of Fe and Mg normally result from infection with malaria-causing *Plasmodia*. Further, the parasites' protein–expression profiles change dramatically. These alterations, when detected and characterized using appropriate and accurate techniques such as trace spectroscopy and (hyper, multi) spectral imaging may be considered as a "fingerprint" of malaria capable of constituting a diagnosis for presence and even severity of malaria (parasitemia).

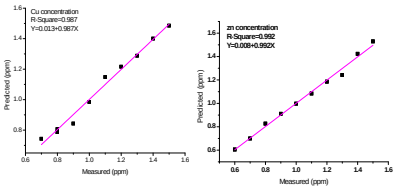

Fig 5. LIBS-PLS calibration models for Cu, Zn based on spectral feature selection.

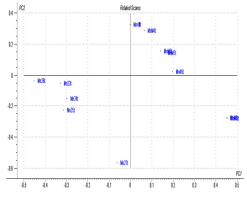

Fig 6. EDXRF PCA score plot for Mn(II) and Mn(IV) in model tissue.

THE RESEARCH PROBLEMS

Accurate and real time trace chemical detection, measurement and modeling via spectroscopy and/spectral imaging which is needed for diagnosis early stages of disease (malaria, cancer) development is a challenge. Example 1: saliva contains analytes in concentrations that are 1000-fold less than those in whole blood. Example 2: Circulating tumour cells (CTCs) in blood, which are the vehicle for cancer metastasizing in other parts of the body, to exist in very low concentrations even in cancer patients.

Although the trace spectroanalytical/imaging tools we use have high versatility, their utility is limited by the complexity of the samples and extreme matrix effects that result in weak spectral signatures; and by the of analysis/mining and interpretation of high-dimensional data. The main challenge is recover the weak signals from the high spectral background and overlapped spectra and images as these are the requisite signatures for detecting and predicting the presence of cancers and malaria especially in their stages.

Malaria confirmation (normally by optical microscopy) is laborious and remains an economic and scientific challenge. Other diagnostic techniques for malarial infection are expensive and knowledge intensive.

Computational intelligence methods that can provide accurate, real time information for clinical purposes are needed as available algorithms have limited prediction quality, numerical instability and slow speed.

MATERIALS & METHODS

We combine the analytical instrumentation tools that we use (LIBS, XRF(EDXRF, TXRF), laser Raman spectromicroscopy, multispectral imaging microscope) with multivariate chemometrics (PLS, ANN, SIMCA, PCA, etc) and machine learning techniques to help reduce the data complexity and increase the information gained. Chemometrics reduces the dimensionality of multivariate data and extracts information which would otherwise be impossible to obtain with classical spectroscopic analysis. Also among other modeling capabilities, chemometrics enables to extract subtle trace signatures from noisy spectra.

Cancer samples are routinely obtained from dogs (metastatic canine cancers share many features with human cancers, e.g. molecular targets, histological appearance, genetics, biological behavior and response to conventional treatments) and from in vitro tissue, body fluid cultures. The dog model samples will soon be drawn from dog tissues specifying the type of cancer (and its stage), sex of dog, age, breed, tissue type, etc since these influence the performance of the multivariate chemometric models.

We use for malaria research *in vitro* cultures of red blood cells infected with *Plasmodium falciparum* prepared as thin smears and without any stain.

Synthetic blood simulates are prepared from analytical grade glycerol, appropriate amounts of poly vinyl chloride (PVC), bovine serum albumin (BSA) and glucose in the range 770-1100 ppm as it occurs in human blood, spiked with multi-element stock solution of elements of interest (Mg, Cu, Zn, Fe) within the range in which they occur in the blood viz Cu: 0.7-1.5 ppm, Zn: 0.6-1.3 ppm, Mg: 15-23, Fe: 0.6-1.6, Se: 93.9-114 µg/l.

In X-ray fluorescence spectrometry samples are irradiated directly and non-invasively by both isotope source and X-ray tube (including TXRF using Ga standard) at different times to obtain fluorescence and scatter spectra (resolution 190 eV) with different SNR.

In the development of novel machine learning techniques (based on multivariate exploratory analysis, clustering, regression, classification) to aid in the analysis and interpretation of the elemental and molecular signatures of measured spectra and images new are being investigating with regard to the performance efficiency and robustness of selected machine learning methods; appropriate cancer biomarkers are being identified.

Multispectral imaging microscopy is accomplished by use of an optical microscope modified by replacing its tungsten light source with a set of light emitting diodes illuminating the samples from different orientations to obtain transmittance, reflectance, and dark-field images by monochrome CMOS camera. Congruent intensity images are recorded at different wavelengths and used to compose multispectral images. Pixels of the multispectral images are used to generate spectral signatures (with the assistance of multivariate chemometrics techniques) to identify infected / non-infected red blood cells.

RESULTS & DISCUSSION

A method for rapid detection and characterization of malaria and parasitemia in thin blood smears on glass, based on presence of hemozoin was developed. The method is based on microscopically imaging red blood cells using different wavelengths of light in UV-NIR wavelength (375-940 nm) region for illumination (multispectral imaging microscopy) .

We have developed a novel new method, energy dispersive X-ray fluorescence and scattering (EDXRFS) spectrometry and demonstrated its utility for rapid non-invasive spectroanalysis protocols for malaria diagnostics using (tiny amounts of blood on Nucleopore paper surface) via multivariate chemometrics modeling; DL of Cu, Zn, Fe, Mg at ppb level (potential forensics application). Chemometrics -LIBS is now being developed.

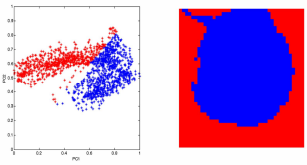

Fig 7. Pixels of a single red blood cell (colour coded-blue) delineated in score space and highlighted in image space.

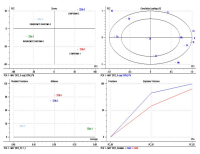

Fig 8. PCA results (scores, x-loadings, correlation loadings (x), influence and explained variance) of trace metals in dog tissue.

TXRF-PCA results of dog tissue show that Cu can be used to identify the presence of rhabdomy sarcoma cancer, Ca for rhabdomy sarcoma 1 type cancer and Br for rhabdomy sarcoma 2 types.

Lymphoma cancer can be diagnosed based on the presence and levels of Se. but Fe and K can be used to determine lymphoma 2 type, while S, P, Zn, Mn are important for lymphoma 1 type. Ca and Zn were found to be present in cancerous tissues but have no utility in cancer diffentiation.

We have developed a chemometrics-assisted EDXRF spectrometry method for trace metal speciation analysis and characterization of Fe, Mn, Cr, Cu in model cancer tissue.

CONCLUSION & FUTURE PLANS

The trace analytical capabilities of spectroscopic and imaging approaches to disease (malaria, cancer) diagnostics are enhanced when combined with multivariate chemometrics and machine learning techniques as this provides greater sensitivity, versatility, multivariate modeling and exploratory capability, and speed.

Multispectral imaging microscopy has been developed and successfully applied to rapid direct malaria diagnostics; it potential to perform diagnostic studies in malaria pathogenesis and biomarker dynamics in blood media at cellular and sub-cellular scale utilizing multivariate spectral imaging and analysis techniques by upgrading the method to hyperspectral imaging and analysis .

We recently acquired laser Raman microimaging spectrometry and look forward to X-ray (confocal micro fluorescence imaging) spectrometry, which has capability of 3-D trace element and speciation analysis for applications that merge optical observation and chemical components analysis/imaging functions, which are necessary for the microanalytical requirements of disease diagnostics at sub-cellular scale and nanomedicine.